Supplementary Material

Methods: Quantitative real-time PCR

DNA was used in a quantitative real-time PCR assay which amplifies codons 72-76 of the *PfCRT* gene on chromosome 7 (Tables S2-S3). TaqMan® probes (Life Technologies) were designed to bind to two different genotypes, one encoding the amino acid sequence CVMNK, which is chloroquine-sensitive (abbreviated CQS), and the other encoding CVIET, which is chloroquine-resistant (abbreviated CQR). Each probe was tagged with a different fluorophore (FAM™ for CVMNK and HEX™ for CVIET) such that samples could be analyzed with both probes simultaneously (multiplexing). Parasite strains 3D7 and Dd2, which have *PfCRT* genotypes CVMNK and CVIET, respectively, were used as positive controls. Samples from Angola were analyzed with a third probe designed to bind to SVMNT (tagged with CAL Fluor® Red 610), using strain HB3 as a positive control. Samples positive for SVMNT (*n* = 13) were excluded from analysis.

Samples were run in triplicate on a Stratagene Mx3000P real-time PCR machine. Control strains (3D7, Dd2, and HB3) were cultured, synchronized to ring stages, and standardized to a density of $2x10^5$ parasites/µl (1). DNA was extracted using QIAamp DNA Mini Kit and serially diluted to densities of $2x10^4$, $2x10^3$, $2x10^2$, and $2x10^1$ genomes/µl for a total of five standards (including $2x10^5$ genomes/µl). The 3D7 and Dd2 standards (as well as HB3 when analyzing samples from Angola) were included in triplicate on each qPCR plate and used to generate standard curves for CVMNK and CVIET, respectively, by plotting C_t values against $\log_{10}$ parasite density. Runs with efficiency below 85% for any genotype were repeated and replaced by runs with higher efficiency. Fluorescence baselines and thresholds were set manually for each sample in MxPro qPCR software (Agilent Technologies). Standard curves were used to convert C_t values to $\log_{10}$ parasite densities. Samples with parasite density below 2 p/µl for any genotype were ruled negative for that genotype; samples negative for both genotypes were excluded from analysis.

Control mixtures of 3D7 and Dd2 at various densities and in different proportions were analyzed to determine the relationships between actual parasite densities and proportions and those determined by qPCR (Fig. S3, S4). The results (polynomial regression relating the observed proportions and parasite densities to actual values) were used to ensure accurate quantification of CQS and CQR parasites in mixed-genotype infections.

Figures

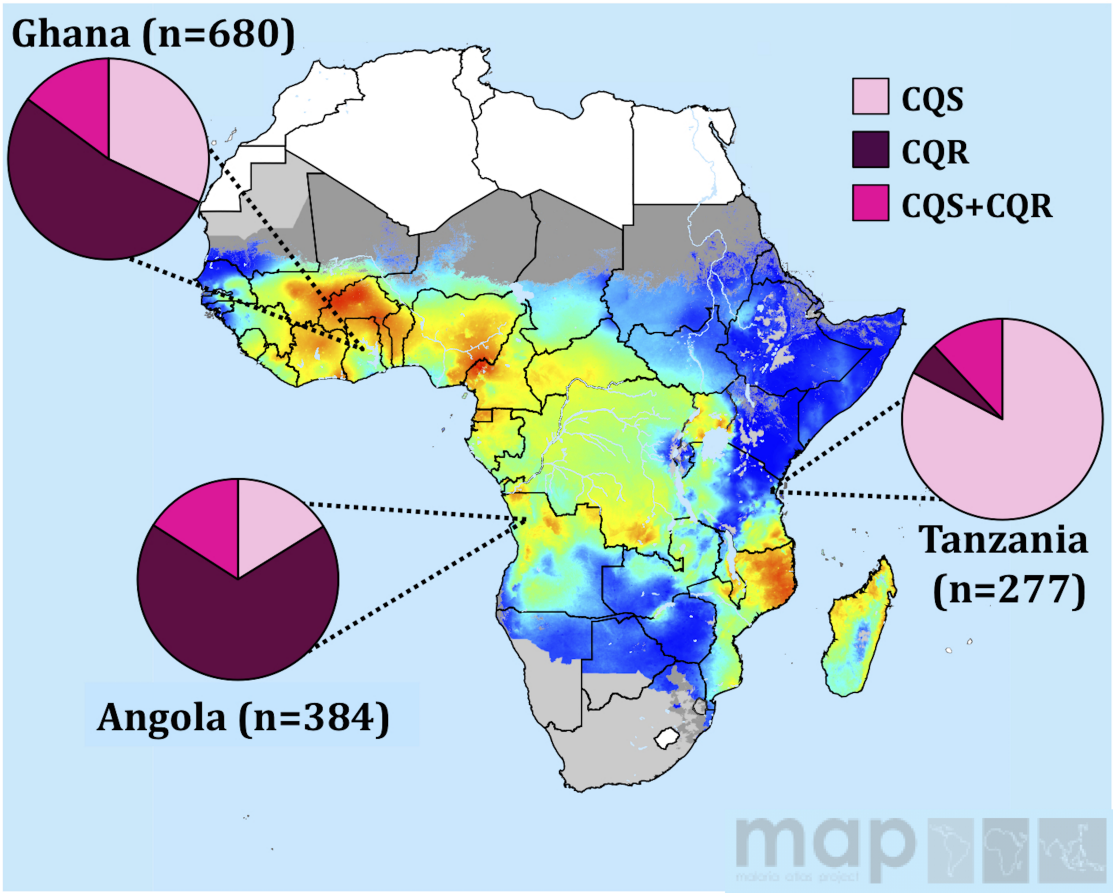

Figure S1. *PfCRT* genotypes of samples from different countries, superimposed on map showing *P. falciparum* endemicity. Pie charts show proportions of infections harboring chloroquine-sensitive (CQS, light pink), chloroquine-resistant (CQR, dark pink), or both alleles (CQS+CQR, medium pink) in samples from Angola, Ghana, and Tanzania. Numbers of samples analyzed from each country are given in parentheses. Shading on the African continent shows endemicity (*Plasmodium falciparum* point prevalence in children ages 2-10 years), ranging from 0% (dark blue) to 70% (red). Map modified from (2).

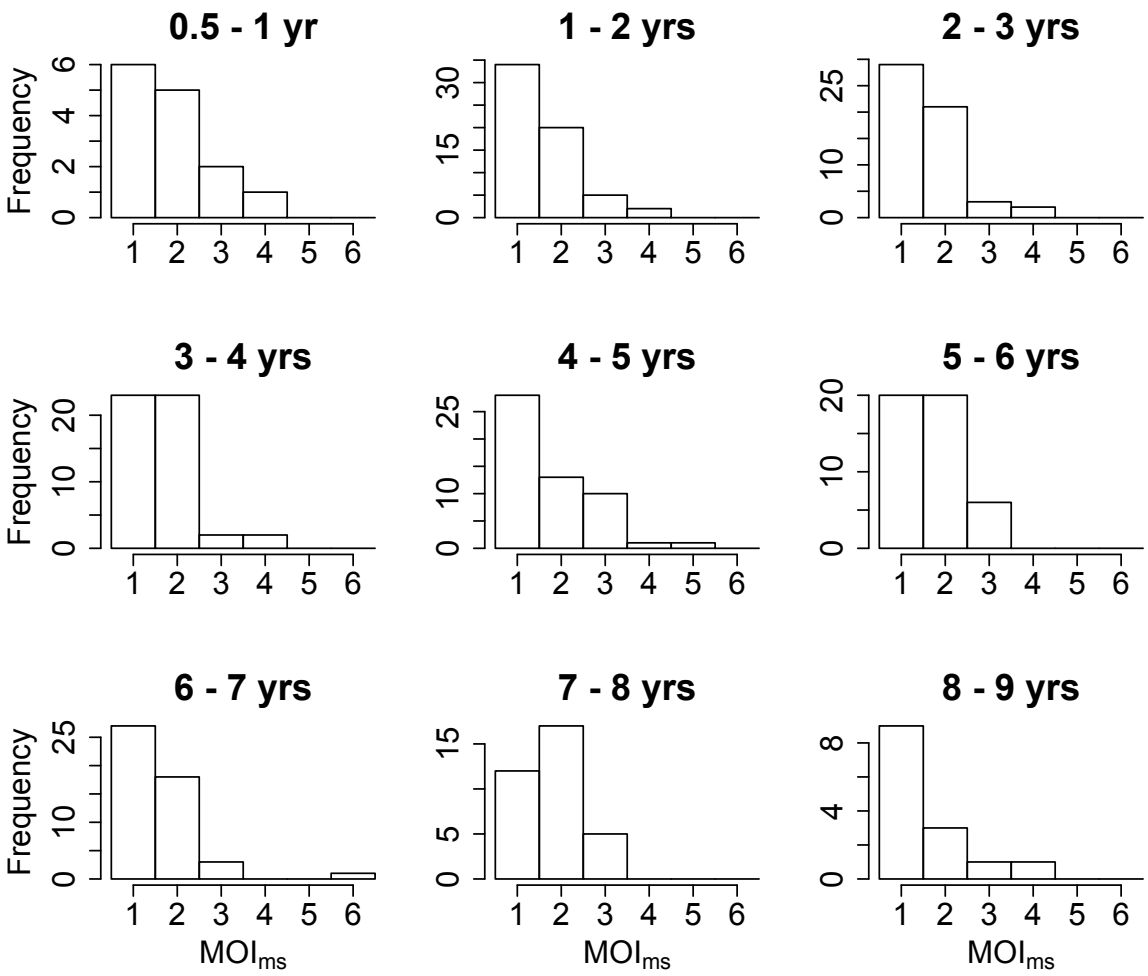

Figure S2. Distribution of MOI_{ms} (number of strains per host) stratified by host age. Each histogram shows the number of hosts in each age group with a given MOI_{ms}.

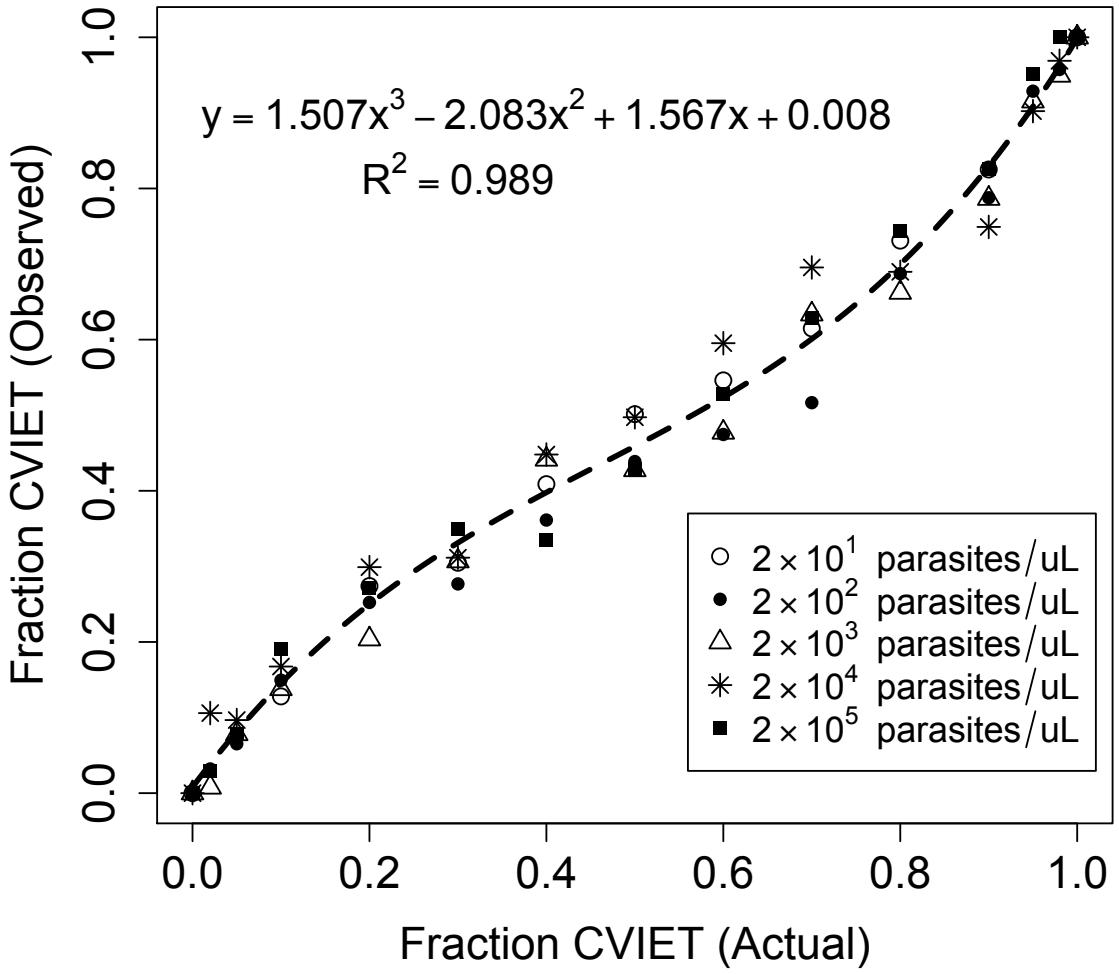

Figure S3. Fraction CVIET (observed) vs. fraction CVIET (actual). The y-axis shows fraction CVIET as determined using qPCR data from control mixtures of 3D7 (CVMNK) and Dd2 (CVIET) at different concentrations (see legend). Dashed line shows polynomial regression (equation on figure).

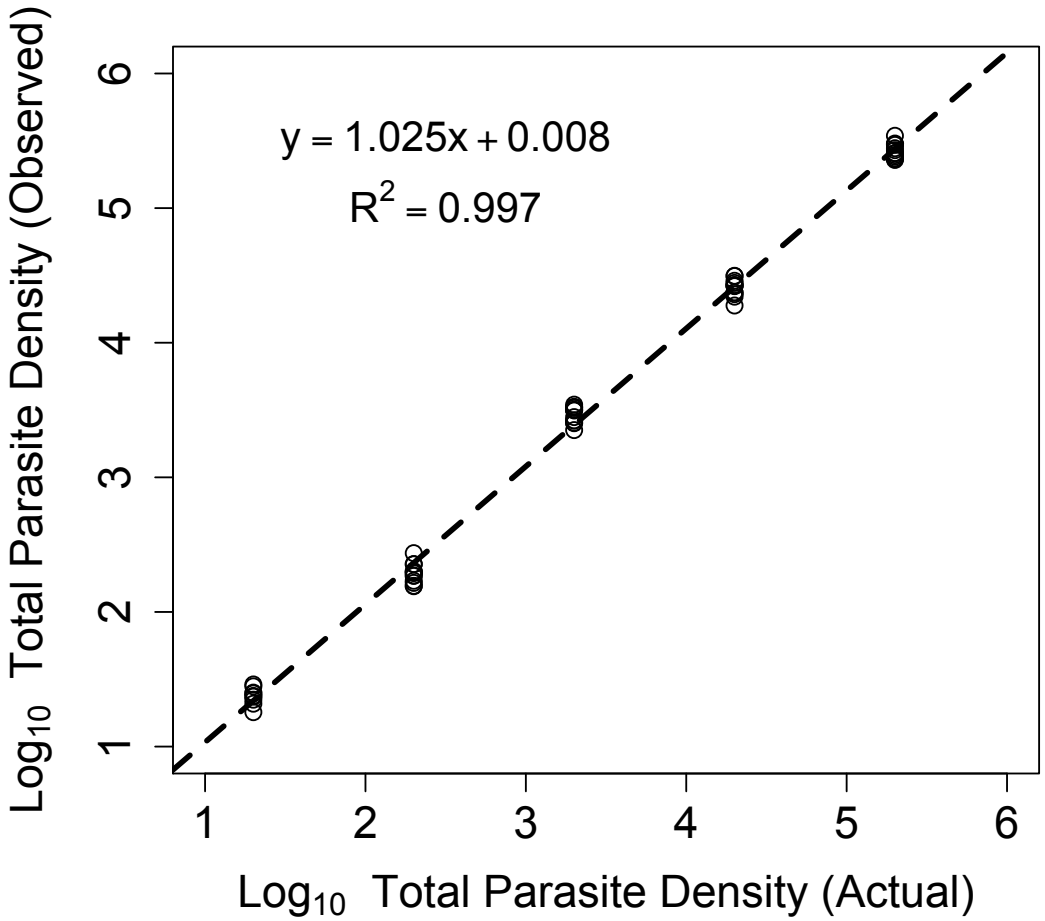

Figure S4. Log_{10} total parasite density: observed values vs. true values. The y-axis shows log_{10} total parasite density as determined using qPCR data from control mixtures of 3D7 and Dd2 (in various proportions; see Fig. S3) at different densities. Dashed line shows linear regression (equation on figure).

Tables

Table S1. Study locations and sample sizes. Numbers in parentheses are samples positive for CVMNK and/or CVIET and – for samples from Angola - negative for SVMNT.

Country	**Study Site**	**# Samples Collected (# Positive by qPCR)**
Angola	Uíge, Uíge Province	196 (190)
	M'banza Congo, Zaire Province	201 (194)
Ghana	Hohoe	199 (177)
	Navrongo	229 (216)
	Sunyani	117 (107)
	Yendi	200 (180)
Tanzania	Bagamoyo District	280 (277)

Table S2. Primer and probe sequences used in *PfCRT* quantitative real-time PCR. Mutant bases in CVIET probe are underlined. FAM™ and HEX™ dyes are owned by Glen Research; CAL Fluor® Red 610 dye and Black Hole Quenchers® (BHQ1 and BHQ2) are owned by Biosearch.

Oligo	**Sequence**
PFCRT-Fwd	5'-TGG TAA ATG TGC TCA TGT GTT T-3'
PFCRT-Rev	5'-AGT TTC GGA TGT TAC AAA ACT ATA GT-3'
CVMNK	5'-FAM-TGT GTA ATG AAT AAA ATT TTT GCT AA-BHQ1-3'
CVIET	5'-HEX-TGT GTA ATT GAA ACA ATT TTT GCT AA-BHQ1-3'
SVMNT	5'-CalRd610-AGT GTA ATG AAT ACA ATT TTT GCT AA-BHQ2-3'

Table S3. *PfCRT* quantitative real-time PCR master mix and thermal cycling protocol.

Master Mix
Invitrogen™ Platinum® Quantitative PCR Supermix-UDG 12.5 µl
$MgCl_2$ (50 mM) 1.25 µl
Fwd primer (10 µM) 0.75 µl
Rev primer (10 µM) 0.75 µl
CVMNK probe (10 µM) 0.25 µl
CVIET probe (10 µM) 0.25 µl
SVMNT probe (10 µM) 0.25 µl
ROX reference dye 0.25 µl
H_2O 3.75 µl
DNA 5 µl
Total volume 25 µl
This reaction can also be run with a total volume of 12.5 µl; in this case all volumes are reduced by half.
Thermal Cycling Protocol
95˚ (6 min.)
[95˚ (15 sec.), 57˚ (1 min.)] x 50
4˚ (hold)

Table S4. Primer sequences for microsatellite loci (3, 4).

Microsatellite	**Primer**	**Sequence**
TA1 Chromosome 6 Nested PCR	Fwd. (primary)	5'-CTA CAT GCC TAA TGA GCA-3'
	Rev.	5'-TTT TAT CTT CAT CCC CAC-3'
	Fwd. (secondary)	5'-HEX-CCG TCA TAA GTG CAG AGC-3'
Polyα Chromosome 4 Nested PCR	Fwd.	5'-AAA ATA TAG ACG AAC AGA-3'
	Rev. (primary)	5'-ATC AGA TAA TTG TTG GTA-3'
	Rev. (secondary)	5'-FAM-GAA ATT ATA ACT CTA CCA-3'
PfPK2 Chromosome 12 Nested PCR	Fwd.	5'-CTT TCA TCG ATA CTA CGA-3'
	Rev. (primary)	5'-CCT CAG ACT GAA ATG CAT-3'
	Rev. (secondary)	5'-HEX-AAA GAA AGG AAC AAG CAG A-3'
TA109 Chromosome 6 Nested PCR	Fwd. (primary)	5'-TAG GGA ACA TCA TAA GGA T-3'
	Rev.	5'-CCT ATA CCA AAC ATG CTA AA-3'
	Fwd. (secondary)	5'-FAM-GGT TAA ATC AGG ACA ACA T-3'
C2M34 Chromosome 2 Non-nested PCR	Fwd.	5'-FAM-TCCCTTTTAAAATAGAAGAAA-3'
	Rev.	5'-GAT TAT ATG AAA GGA TAC ATG-3'
C3M69 Chromosome 3 Non-nested PCR	Fwd.	5'-HEX-AAT AGG AAC AAA TCA TAT TG-3'
	Rev.	5'-AGA TAT CCA GGT AAT AAA AAG-3'

Table S5. PCR master mixes and thermal cycling protocols for amplifying neutral microsatellites**.** Protocols for non-nested PCRs are from (5) and for nested PCRs are from (6).

Non-nested PCR: Master Mix Promega PCR Master Mix 7.5 µl H_2O 4.3 µl Fwd primer (stock conc. 10 µM) 0.6 µl Rev primer (stock conc. 10 µM) 0.6 µl DNA 2 µl Total volume 15 µl
Non-nested PCR: Thermal Cycling Protocol 94˚ (2 min.) [94˚ (30 sec.), 50˚ (30 sec.), 60˚ (30 sec.)] x 5 [94˚ (30 sec.), 45˚ (30 sec.), 60˚ (30 sec.)] x 40 4˚ (hold)
Nested PCR Master Mix (Primary and Secondary Reactions) Promega PCR Master Mix 7.5 µl H_2O 5.3 µl Fwd primer (stock conc. 10 µM) 0.6 µl Rev primer (stock conc. 10 µM) 0.6 µl DNA 1 µl Total volume 15 µl
Nested PCR Thermal Cycling Protocol: Primary Reaction 94˚ (2 min.) [94˚ (30 sec.), 42˚ (30 sec.), 40˚ (30 sec.), 65˚ (40 sec.)] x 25 65˚ (2 min.) 4˚ (hold)
Nested PCR Thermal Cycling Protocol: Secondary Reaction 94˚ (2 min.) [94˚ (20 sec.), 45˚* (20 sec.), 65˚ (30 sec.)] x 25 65˚ (2 min.) 4˚ (hold) *For TA109 change 45˚ to 59˚

Table S6. Distributions of multiplicity of infection (MOI) for Angola (based on microsatellite genotyping) and Tanzania (based on Carlsson et al. (7)).

	MOI=1	MOI=2	MOI=3	MOI=4	MOI=5	MOI=6
Frequency (Angola)	0.50	0.36	0.10	0.03	0.003	0.003
Frequency (Tanzania)	0.22	0.40	0.24	0.08	0.04	0.02

Table S7. Number and frequency of infections positive for CQS and/or CQR alleles in Angola, Ghana, and Tanzania. Prevalence of mixed-genotype infections (CQS+CQR) did not vary significantly between countries (X^2 = 2.898, d.f. = 2, p = 0.235) but the frequencies of CQS and CQR varied markedly (X^2 = 326.974, d.f. = 2, p = 9.967×10^{-72}).

	Angola (*n*)	**Ghana (*n)***	**Tanzania (*n*)**	**All (*n*)**
CQS only	65 (16.9%)	218 (32.1%)	229 (82.7%)	512 (38.2%)
CQR only	255 (66.4%)	361 (53.1%)	15 (5.4%)	631 (47.1%)
CQS + CQR	64 (16.7%)	101 (14.9%)	33 (11.9%)	198 (14.8%)
Total	384	680	277	1,341

Structure and implementation of null model of CQS+CQR mixed-genotype infections

The data required for the null model are CQS and CQR allele frequencies and the distribution of multiplicity of infection (MOI) in the population. Using these data, it is possible to estimate the probability of an infection with any given composition (for example, the probability that a host will have three strains, of which one is CQR and two are CQS). With these probabilities, and working under the null assumption that the numbers of CQS and CQR parasites in a host will be strictly proportional to the numbers of strains of each genotype, the model is used to calculate the expected mean proportion of CQR parasites in all CQS+CQR mixed infections in a population.

CQS and CQR allele frequencies were estimated from single-genotype samples used in our study. Data on multiplicity of infection (MOI) for Angola were also collected in our study, and data on MOI for Tanzania were obtained from contemporaneous work on the within-host diversity of malaria infections in Bagamoyo District, Tanzania (7). From the published data, the estimation of MOI was similar to the method we used for Angola: MOI = maximum number of alleles observed at *msp1* or *msp2*. Given the distributions of MOI and CQS/CQR allele frequencies, we can calculate the expected proportion CQR in CQS+CQR mixed infections for each population, as follows. Let m be the maximum MOI observed

in the population, f_i the frequency of infections with MOI=*i, p* the frequency of the CQR allele and *q* the frequency of the CQS allele. The expected proportion CQR is obtained by summing over all possible values multiplied by their probabilities. For an infection with *i* total strains, of which *j* are resistant, the expected proportion CQR (under the null model) is *j/i;* the rest of the summand gives the probability of such an infection. The denominator is the total probability of all mixed-genotype infections; dividing by this ensures that the estimated probabilities sum to 1.

$$\frac{\sum_{i=2}^{m}\sum_{j=1}^{i-1} f_i\left(\frac{j}{i}\right)\binom{i}{j}p^j q^{i-j}}{\sum_{i=2}^{m}\sum_{j=1}^{i-1} f_i\binom{i}{j}p^j q^{i-j}}$$

Both Angola and Tanzania were found to have MOI ranging from 1 to 6 (frequencies shown in Table S6), and our estimate of the CQR allele frequencies in Tanzania and Angola were 0.06 and 0.80, respectively. Calculating the denominator of the formula above gives the expected proportion of all infections expected to have both CQS and CQR genotypes; this value comes out to be 0.122 for Tanzania and 0.185 for Angola, which are not statistically different from the observed frequencies of mixed-genotype infections in our sample (0.119 for Tanzania, p=0.96; 0.169 for Angola, p=0.25). Therefore, the MOI distributions derived from Carlsson et al. and from our microsatellite data are consistent with the observed frequencies of CQS+CQR mixed-genotype infections. The formula above gives the expected proportion CQR in mixed-genotype infections, which is found to be 0.544 for Angola and 0.377 for Tanzania. For each of these countries, the observed average proportion CQR is significantly lower than expected (one-sample, two-sided *t*-tests of logit-transformed proportion data (p = 0.01 for Angola, $p = 9\text{x}10^{-5}$ for Tanzania).

Information on MOI was not available for Ghana, but it is possible to test against a simplified null model, which makes no assumptions about the distribution of MOI in the population. Since the frequency of the CQR allele in Ghana was 0.62, a neutral model would predict the proportion CQR to be greater than or equal to 0.5. A one-sample, two-sided *t*-test of logit-transformed proportion data (two-sided to be conservative, although a one-sided test would also be appropriate) shows that the average proportion CQS is significantly less than 0.5 ($p=4\text{x}10^{-8}$). The rejection of the simplified null model, in this case, also allows us to reject the full null model. Adding information about the distribution of MOI will always yield an expected mean between 0.5 and the CQR allele frequency (0.62). If the average proportion CQR in Ghana is significantly less than 0.5, it will also be significantly less than any value greater than 0.5; therefore, we know that the prediction of the more sophisticated null model will also be rejected.

The null model can be modified to incorporate the differences observed between CQS and CQR parasites in single-genotype infections. If the average density of CQR in single infections is some fraction *x* of the average density of CQS in single infections, then the expected proportion CQR in a host with *j* resistant and *i-j* sensitive strains becomes $(xj)\div(i+j(x-1))$ and the rest of the model is unchanged.

The primary limitation of this null model stems from the fact that methods such as microsatellite genotyping will tend to underestimate MOI. As a result, the proportions predicted by this null model will tend to err toward 0.5, regardless of whether this is above or below the predicted value. With this in mind, this model provides estimates that fall in between two extremes, the first being 0.5 (the proportion CQR expected if all CQS+CQR mixed infections comprise one strain of each genotype) and the second being the frequency of the CQR allele in the population (which is the limit approached by the model as multiplicity of infection goes to infinity).

References

1. Methods Manual for Product Testing of Malaria Rapid Diagnostic Tests (v. 5) Foundation for Innovative New Diagnostics, World Health Organization (Global Malaria Programme), Centers for Disease Control and Prevention (Division of Parasitic Diseases, Malaria Branch), 2012.
2. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malar J. 2011;10:378.
3. Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP. Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology. 1999;119 (Pt 2):113-25.
4. McCollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA. Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrob Agents Chemother. 2007;51(6):2085-91.
5. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol Biol Evol. 2003;20(9):1526-36.
6. Roper C, Pearce R, Bredenkamp B, Gumede J, Drakeley C, Mosha F, et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. Lancet. 2003;361(9364):1174-81.
7. Carlsson AM, Ngasala BE, Dahlstrom S, Membi C, Veiga IM, Rombo L, et al. Plasmodium falciparum population dynamics during the early phase of anti-malarial drug treatment in Tanzanian children with acute uncomplicated malaria. Malar J. 2011;10:380.